International Journal of Approximate Reasoning 52 (2011) 1312–1324

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Possibilistic analysis of arity-monotonic aggregation operators and its relation to bibliometric impact assessment of individuals

Marek Gągolewski *, Przemysław Grzegorzewski

Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland
Faculty of Mathematics and Information Science, Warsaw University of Technology, pl. Politechniki 1, 00-661 Warsaw, Poland

ARTICLE INFO

Article history:
Available online 15 February 2011

Keywords:
Aggregation operators
Possibility theory
S-statistics
h-Index
OWMax

ABSTRACT

A class of arity-monotonic aggregation operators, called impact functions, is proposed. This family of operators forms a theoretical framework for the so-called Producer Assessment Problem, which includes the scientometric task of fair and objective assessment of scientists using the number of citations received by their publications.

The impact function output values are analyzed under right-censored and dynamically changing input data. The qualitative possibilistic approach is used to describe this kind of uncertainty. It leads to intuitive graphical interpretations and may be easily applied for practical purposes.

The discourse is illustrated by a family of aggregation operators generalizing the well-known Ordered Weighted Maximum (OWMax) and the Hirsch *h*-index.

1. Introduction

In many areas of human activity like engineering, science, statistics, economy or social sciences, data summarization is often required for further reasoning and decision making. A kind of synthesis of all individual inputs can be achieved by an appropriate aggregation.

Aggregation operators merge several numerical values into a single, representative one. Thus, from the perspective of mathematics, aggregation operators are just projections from a multidimensional state space into a single dimension. Apart from particular applications, the theory of aggregation operators is a rapidly developing mathematical domain (we refer the reader to [1] for the recent state of the art monograph).

Most often, aggregation operators are considered for a fixed number of arguments. This might be too restrictive in some applications. We face such situation in the so-called Producer Assessment Problem (PAP) described in Section 3, where given alternatives are rated not only with respect to the quality of delivered items but also to their productivity. The issue of fair assessment of scientists based on the number of citations gained by their papers is the most representative instance of the PAP. The *h*-index proposed by Hirsch [2] is one of the most widely known tools used in this domain. This is the reason why aggregation operators defined for arbitrary number of arguments are of interest.

The paper is organized as follows. In Section 2 we present the conventional notation used throughout the article and recall some general information on aggregation functions and possibility measures. In Section 3 we describe the class of impact functions which form a model for the Producer Assessment Problem. Also we present some basic properties of impact functions.

* Corresponding author at: Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland. Tel.: +48 22 38 10 393; fax: +48 22 38 10 105.
E-mail addresses: gagolews@ibspan.waw.pl (M. Gągolewski), pgrzeg@ibspan.waw.pl (P. Grzegorzewski).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2011.01.010

Then, in Section 4, we study a class of aggregation operators, called the S-statistics [3,4]. These functions generalize the Ordered Weighted Maximum (OWMax) and the h-index.

In Section 5 we analyze the behavior of impact functions from a dynamic perspective, in which two possible situations are taken into consideration. Firstly, some of the elements of the input vector may be likely to increase their values, e.g. when papers gain more citations. Secondly, we may also be faced with right-censored data, e.g. when we gather bibliometric records from databases that do not cover the whole spectrum of journals, i.e. that underestimate the true number of citations. In both cases we are interested in predicting the effects of input vector alteration on the impact function's output value. Here we utilize possibility theory to express this kind of uncertainty. Our results are then illustrated in Section 6. Finally, Section 7 summarizes the paper.

2. Preliminaries

2.1. Basic notation

Let $\mathbb{I} = [a, b]$ denote any nonempty closed interval of extended real numbers $\overline{\mathbb{R}} = [-\infty, \infty]$. The family of all subsets of $\mathbb{I}$ will be denoted by $\mathfrak{P}(\mathbb{I})$. Unless stated otherwise, $n, m \in \mathbb{N}$. We assume that $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ denotes the set of all nonnegative integers while $[n] = \{1, 2, \ldots, n\}$.

Given any $\mathbf{x} = (x_1, \ldots, x_n),\ \mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{I}^n$, we write $\mathbf{x} \leqslant \mathbf{y}$ iff $(\forall i \in [n])\ x_i \leqslant y_i$. Similarly, we say that $\mathbf{x} < \mathbf{y}$ when $\mathbf{x} \leqslant \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$. Additionally, $\mathbf{x} \cong \mathbf{y}$ iff there exists a permutation σ of $[n]$ such that $(x_1, \ldots, x_n) = (y_{\sigma(1)}, \ldots, y_{\sigma(n)})$.

Let $x_{(i)}$ denote the ith-smallest value of $\mathbf{x} = (x_1, \ldots, x_n)$ and $(n * x)$ the n-tuple $(x, x, \ldots, x) \in \mathbb{I}^n$.

The set of all vectors of arbitrary length with elements in $\mathbb{I}$, i.e. $\bigcup_{n=1}^{\infty} \mathbb{I}^n$, will be denoted by $\mathbb{I}^{1,2,\ldots}$. For any $\mathbf{x} \in \mathbb{I}^n, \mathbf{y} \in \mathbb{I}^m$ and any function g defined on $\mathbb{I}^{n+m}$ the notation $\mathrm{g}(\mathbf{x}, \mathbf{y})$ stands for $\mathrm{g}(x_1, \ldots, x_n, y_1, \ldots, y_m)$.

2.2. Aggregation functions

To establish a point of reference for further discussion, let us first recall the notion of the aggregation function extended to any number of arguments. Here is the definition given in [1]. Note that much more restrictive sine qua non conditions were proposed by Calvo et al. [5,6] by means of so-called α- and β-orderings.

From now on, let $\mathcal{E}(\mathbb{I})$ denote the family of all *aggregation operators* in $\mathbb{I}^{1,2,\ldots}$, i.e. all the functions from $\mathbb{I}^{1,2,\ldots}$ to $\overline{\mathbb{R}}$. This description reflects the very general idea of data summarization/synthesis mentioned in the Introduction.

Definition 1. An (**extended**) **aggregation function** in $\mathbb{I}^{1,2,\ldots}$ is any function $\mathsf{A} \in \mathcal{E}(\mathbb{I})$ such that for any n

(A1) is nondecreasing in each variable, i.e. $(\forall \mathbf{x}, \mathbf{y} \in \mathbb{I}^n)\mathbf{x} \leqslant \mathbf{y} \Rightarrow \mathsf{A}(\mathbf{x}) \leqslant \mathsf{A}(\mathbf{y})$,
(A2) fulfills the lower boundary condition: $\inf_{\mathbf{x} \in \mathbb{I}^n} \mathsf{A}(\mathbf{x}) = a$,
(A3) fulfills the upper boundary condition: $\sup_{\mathbf{x} \in \mathbb{I}^n} \mathsf{A}(\mathbf{x}) = b$.

Typical examples of aggregation functions are: the sample minimum, maximum, arithmetic mean, median, Bonferroni mean [7], OWA [8] and OWMax [9] operators. On the other hand, the sample size, sum or any constant function generally are not aggregation functions in the sense of Definition 1.

It is worth noting that axioms (A1) and (A2) imply $\mathsf{A}(n * a) = a$. We also have $\mathsf{A}(n * b) = b$ by (A1) and (A3).

The set of all extended aggregation functions in $\mathbb{I}^{1,2,\ldots}$ will be denoted by $\mathcal{E}_{\mathcal{A}}(\mathbb{I})$.

2.3. Possibility theory

Let us recall here some definitions and concepts relevant to fuzzy measures that will be useful in further considerations.

Definition 2. A function $\mu : \mathfrak{P}(\mathbb{I}) \to [0, 1]$ is a **fuzzy measure** (also called *normalized capacity) if*

(a) $\mu(\emptyset) = 0$ and $\mu(\mathbb{I}) = 1$ (normalization),
(b) for all $A, B \in \mathfrak{P}(\mathbb{I})$, if $A \subseteq B$, then $\mu(A) \leqslant \mu(B)$ (monotonicity).

Possibility theory is one of the formal representations of uncertainty. It may be used to describe the knowledge of an agent about the value of some quantity ranging on $\mathbb{I}$; some states may be marked on a scale with "impossible" state on one side and "normal" or "unsurprising" on the other.

Definition 3. A fuzzy measure Pos is called a **possibility measure** iff for any family $\{A_k \in \mathfrak{P}(\mathbb{I}) : k \in K\}$ and an arbitrary index set K,

$$\mathsf{Pos}\left(\bigcup_{k \in K} A_k\right) = \sup_{k \in K} \mathsf{Pos}(A_k). \tag{1}$$

With each possibility measure Pos we may associate another fuzzy measure Nec, called **necessity measure**, defined by

$$\mathsf{Nec}(A) := 1 - \mathsf{Pos}(\bar{A}), \tag{2}$$

where $A \in \mathfrak{P}(\mathbb{I})$ and $\bar{A}$ denotes the complement of A. It may be shown that for every $A \in \mathfrak{P}(\mathbb{I})$ and for any possibility measure Pos with the associated necessity measure Nec the following relations hold:

(a) $\mathsf{Nec}(A) > 0 \Rightarrow \mathsf{Pos}(A) = 1$,
(b) $\mathsf{Pos}(A) < 1 \Rightarrow \mathsf{Nec}(A) = 0$.

Moreover, every possibility measure Pos is uniquely determined by a **possibility distribution** $\pi : \mathbb{I} \to [0, 1]$, i.e. for any $A \in \mathfrak{P}(\mathbb{I})$

$$\mathsf{Pos}(A) = \sup_{x \in A} \pi(x). \tag{3}$$

The possibility distribution plays a central role in possibility theory. It represents the knowledge distinguishing what is plausible from what is surprising, e.g. $\pi(x) = 0$ means that x is considered impossible, while $\pi(x) = 1$ means that x is totally possible. For more details concerning possibility theory, further references and other representations of uncertainty the reader is referred, e.g. to [10–14].

3. Impact functions and their properties

3.1. The Producer Assessment Problem

Consider a **producer** (e.g. a writer, scientist, artist, craftsman) and a nonempty set of his **products** (e.g. books, papers, works, goods). Suppose that each product is given a **rating** (of quality, popularity, etc.) which is a single number in $\mathbb{I} = [a, b]$, where a denotes the lowest admissible valuation. Some typical examples of such situation are listed in Table 1.

It is clear that each possible state of producer's activity can be described by a point in $\mathbb{I}^{1,2,\dots}$. The **Producer Assessment Problem** (or PAP for short) involves constructing and analyzing aggregation operators which can be used for rating producers [4]. A family of such functions should take into account the two following aspects of producer's quality:

- the ability to make highly-rated products,
- overall productivity.

Clearly, the first component can be properly described by a very broad class of (extended) aggregation functions (Definition 1). However, note that any two restrictions $\mathsf{A}|_{\mathbb{I}^n}$ and $\mathsf{A}|_{\mathbb{I}^m}$ of an extended aggregation function A, where $n \neq m$, are generally not necessarily related. Formally, we say that the family $\mathcal{E}_{\mathcal{A}}(\mathbb{I})$ is **arity-free**, i.e. $(\forall n \neq m)$ $(\forall \mathsf{f}^{(n)} \in \mathcal{E}_{\mathcal{A}}(\mathbb{I})|_{\mathbb{I}^n})$ it holds

$$\left\{\mathsf{A}|_{\mathbb{I}^m} : \mathsf{A} \in \mathcal{E}_{\mathcal{A}}(\mathbb{I}),\ \mathsf{A}|_{\mathbb{I}^n} = \mathsf{f}^{(n)}\right\} = \mathcal{E}_{\mathcal{A}}(\mathbb{I})|_{\mathbb{I}^m}.$$

As stated above, in practice we are also interested in distinguishing between entities of different productivities. Therefore, we need some sine qua non requirements which should be satisfied by such assessing functions.

3.2. Impact functions

Consider the following definition. Its slightly modified version was given in [4].

Definition 4. An **impact function** in $\mathbb{I}^{1,2,\dots}$ is a function $\mathsf{J} \in \mathcal{E}(\mathbb{I}),\ \mathbb{I} = [a, b]$, which

(I1) is nondecreasing in each variable: $(\forall n)$ $(\forall \mathbf{x}, \mathbf{y} \in \mathbb{I}^n) \mathbf{x} \leqslant \mathbf{y} \Rightarrow \mathsf{J}(\mathbf{x}) \leqslant \mathsf{J}(\mathbf{y})$,
(I2) fulfills the weak lower boundary condition: $\inf_{\mathbf{x} \in \mathbb{I}^{1,2,\dots}} \mathsf{J}(\mathbf{x}) = a$,
(I3) fulfills the weak upper boundary condition: $\sup_{\mathbf{x} \in \mathbb{I}^{1,2,\dots}} \mathsf{J}(\mathbf{x}) = b$,

Table 1
The Producer Assessment Problem – typical examples.

	Producer	Products	Rating method	Discipline
A	Scientist	Scientific articles	Number of citations	Scientometrics
B	Scientific institute	Scientists	The *h*-index	Scientometrics
C	Web server	Web pages	Number of in-links	Webometrics
D	Artist	Paintings	Auction price	Auctions
E	Billboard company	Advertisements	Sale results	Marketing

(I4) is arity-monotonic, i.e. $(\forall n, m)\ (\forall \mathbf{x} \in \mathbb{I}^n)\ (\forall \mathbf{y} \in \mathbb{I}^m)\ \mathsf{J}(\mathbf{x}) \leqslant \mathsf{J}(\mathbf{x}, \mathbf{y})$,
(I5) is symmetric, i.e. $(\forall n)\ (\forall \mathbf{x}, \mathbf{y} \in \mathbb{I}^n) \mathbf{x} \cong \mathbf{y} \Rightarrow \mathsf{J}(\mathbf{x}) = \mathsf{J}(\mathbf{y})$.

Conditions (I1) and (I4) correspond to the principle called "the more the better", which is justified in many practical instances of the PAP. According to (I5) each product is of equal importance and the overall rating is not affected by the presentation order of the products.

Such formal model given for the PAP allows us to abstract from its context-dependent interpretation (to avoid any bias) and focus solely on the analysis of its mathematical properties.

The family of all impact functions will be denoted by $\mathcal{E}_{\mathcal{I}}(\mathbb{I})$. Note that the set of requirements given in Definition 4 is similar to the axiomatization proposed by Woeginger [15,16] for the so-called bibliometric impact indices (for other axiomatizations see [17–20]).

It should be stressed that impact functions are not necessarily extended aggregation functions (in the sense of Definition 1), because axioms (A2) and (A3) are replaced by their weaker forms (I2), (I3), respectively. For example, (A3) in conjunction with (I1), (I4), (I5) yields $(\forall \mathbf{x} \in \mathbb{I}^n)\ (\exists i \in [n]) x_i = b \Rightarrow \mathsf{J}(x_1, \ldots, x_n) = b$, which generally is not a desirable property.

Moreover, (I1), (I4) and the closedness of $\mathbb{I}$ imply that (I2) is equivalent to the condition $\mathsf{J}(a) = a$, and (I3) holds iff $\lim_{n\to\infty} \mathsf{J}(n * b) = b$.

3.3. Alternative definition

From now on, let $P_{(\mathrm{I1})}$ stand for the family of nondecreasing functions (i.e. functions satisfying axiom (I1) in Definition 4), $P_{(\mathrm{I2})}$ denote the functions fulfilling (I2), and so on. For example, $P_{(\mathrm{A1})} = P_{(\mathrm{I1})}$, $P_{(\mathrm{A2})} \subseteq P_{(\mathrm{I2})}$, $P_{(\mathrm{A3})} \subseteq P_{(\mathrm{I3})}$ and $\mathcal{E}_{\mathcal{I}}(\mathbb{I}) = P_{(\mathrm{I1})} \cap P_{(\mathrm{I2})} \cap \cdots \cap P_{(\mathrm{I5})}$.

Consider the following relation $\trianglelefteq$ on $\mathbb{I}^{1,2,\ldots}$ which will be needed in further discussions. For any $\mathbf{x} \in \mathbb{I}^n$ and $\mathbf{y} \in \mathbb{I}^m$

$$\mathbf{x} \trianglelefteq \mathbf{y} \iff n \leqslant m \text{ and } x_{(n-i+1)} \leqslant y_{(m-i+1)} \text{ for all } i \in [n]. \tag{4}$$

Of course, $\trianglelefteq$ is a partial order. Moreover, we write $\mathbf{x} \triangleleft \mathbf{y}$ if $\mathbf{x} \trianglelefteq \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$.

The following result is very important when considering the class of impact functions.

Theorem 5. *Let* $\mathsf{F} \in \mathcal{E}(\mathbb{I})$. *Then* $\mathsf{F} \in P_{(\mathrm{I1})} \cap P_{(\mathrm{I4})} \cap P_{(\mathrm{I5})}$ *if and only if* $(\forall \mathbf{x}, \mathbf{y})\ \mathbf{x} \trianglelefteq \mathbf{y} \Rightarrow \mathsf{F}(\mathbf{x}) \leqslant \mathsf{F}(\mathbf{y})$.

Proof. $(\Rightarrow)$ Let $\mathsf{F} \in P_{(\mathrm{I1})} \cap P_{(\mathrm{I4})} \cap P_{(\mathrm{I5})}$ and take any $\mathbf{x} \trianglelefteq \mathbf{y},\ \mathbf{x} \in \mathbb{I}^n,\ \mathbf{y} \in \mathbb{I}^m$. We consider two cases.

(a) $n = m$. As $(\forall i \in [n])\ x_{(i)} \leqslant y_{(i)}$, then by (I1) and (I5) it holds $\mathsf{F}(\mathbf{x}) \leqslant \mathsf{F}(\mathbf{y})$.
(b) $n < m$. Let $\mathbf{y}' = (y_{(m)}, y_{(m-1)}, \ldots, y_{(m-n+1)}) \in \mathbb{I}^n$. By (I1) and (I5) we have $\mathsf{F}(\mathbf{x}) \leqslant \mathsf{F}(\mathbf{y}')$ and by (I4) $\mathsf{F}(\mathbf{y}') \leqslant \mathsf{F}(\mathbf{y})$.

$(\Leftarrow)$ Take any $\sigma, \sigma' \in \mathfrak{S}_{[n]}$ and $\mathbf{x} \in \mathbb{I}^n$. We have $(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \trianglelefteq (x_{\sigma'(1)}, \ldots, (x_{\sigma'(n)})$ and $(x_{\sigma'(1)}, \ldots, x_{\sigma'(n)}) \trianglelefteq (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. Therefore $\mathsf{F}(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = \mathsf{F}(x_{\sigma'(1)}, \ldots, (x_{\sigma'(n)})$ and hence F is symmetric (axiom (I5)).

Now, take any $\mathbf{x}, \mathbf{y} \in \mathbb{I}^n$ such that $\mathbf{x} \leqslant \mathbf{y}$. As $\mathbf{x} \leqslant \mathbf{y}$ implies that $\mathbf{x} \trianglelefteq \mathbf{y}$, we have $\mathsf{F}(\mathbf{x}) \leqslant \mathsf{F}(\mathbf{y})$ and so F is nondecreasing in each variable (axiom (I1)).

Lastly, for all $\mathbf{x} \in \mathbb{I}^n$ and any $\mathbf{y} \in \mathbb{I}^m$, $(\mathbf{x}) \trianglelefteq (\mathbf{x}, \mathbf{y})$, hence $\mathsf{F}(\mathbf{x}) \leqslant \mathsf{F}(\mathbf{x}, \mathbf{y})$. Therefore F is arity-monotonic (axiom (I4)) and the proof is complete. □

Therefore, the class of impact functions is equivalent to the set of order-preserving maps (morphisms) from the partially ordered set $(\mathbb{I}^{1,2,\ldots}, \trianglelefteq)$ to $(\mathbb{R}, \leqslant)$ that fulfill (I2) and (I3).

3.4. Basic properties

Axiomatic modeling in decision making dates back to de Finetti [21], von Neumann and Morgenstern [22], Arrow [23] (impossibility theorem in the problem of social states ordering) and May [24] (group decision functions).

Formally, a **property** P of functions in $\mathcal{E}(\mathbb{I})$ is an appropriate subset of $\mathcal{E}(\mathbb{I})$.

Below we discuss some basic properties which may be useful to describe the behavioral aspects of aggregation operators. Some of them are desirable in particular instances of the PAP.

It can sometimes be justifiable to treat $a = \min \mathbb{I}$ as the "minimal **admissible** quality". Then adding new products with such rating should not affect the overall ranking (such elements are treated as negligible). By formalizing this idea we obtain the following property.

Definition 6. We say that a function $\mathsf{F} \in P_{(\mathrm{I2})} \cap P_{(\mathrm{I3})}$ is a**-insensitive** iff $\mathsf{F}(\mathbf{x}, a) = \mathsf{F}(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{I}^{1,2,\ldots}$.

The family of a-insensitive functions will be denoted $P_{(\mathrm{a0})}$. By Definition 4 we get

$$P_{(\mathrm{a0})} \cap P_{(\mathrm{I1})} \subseteq P_{(\mathrm{I4})} \cap P_{(\mathrm{I1})}.$$

The above property may be strengthened as follows.

Definition 7. We say that a function $\mathsf{F} \in P_{(\mathrm{I2})} \cap P_{(\mathrm{I3})}$ is F-**insensitive** iff $(\forall \mathbf{x} \in \mathbb{I}^{1,2,\ldots})$ $(\forall y \in \mathbb{I}) y \leqslant \mathsf{F}(\mathbf{x}) \Rightarrow \mathsf{F}(\mathbf{x}, y) = \mathsf{F}(\mathbf{x})$.

The family of F-insensitive aggregation operators will be denoted by $P_{(\mathrm{F0})}$. Note that if $\mathsf{F} \in \mathcal{E}_{\mathcal{I}}(\mathbb{I})$ then $\mathsf{F} \in P_{(\mathrm{F0})}$ iff $(\forall \mathbf{x} \in \mathbb{I}^{1,2,\ldots})\mathsf{F}(\mathbf{x}, \mathsf{F}(\mathbf{x})) = \mathsf{F}(\mathbf{x})$. We may also see that $P_{(\mathrm{F0})} \subseteq P_{(\mathrm{a0})}$.

On the other hand, in some applications it would be admissible that the function is sensitive for the addition of elements greater that $\mathsf{F}(\mathbf{x})$.

Definition 8. We say that a function $\mathsf{F} \in P_{(\mathrm{I2})} \cap P_{(\mathrm{I3})}$ is F+**sensitive** iff $(\forall \mathbf{x} \in \mathbb{I}^{1,2,\ldots})$ $(\forall y \in \mathbb{I})$ $y > \mathsf{F}(\mathbf{x}) \Rightarrow \mathsf{F}(\mathbf{x}, y) > \mathsf{F}(\mathbf{x})$.

The family of F+sensitive functions will be denoted by $P_{(\mathrm{F+})}$.

For certain aggregation operators such a vector $\mathbf{x}$ may exist that any increment of the values of its coordinates never results in a change of the aggregator's output. We say then that such operator is *saturated* at $\mathbf{x}$. In this case only the addition (concatenation) of elements to $\mathbf{x}$ may increase the output value. Here is a formal definition of a saturable operator.

Definition 9. We say that a function $\mathsf{F} \in P_{(\mathrm{I1})} \cap P_{(\mathrm{I4})}$ is **saturable** (denoted $\mathsf{F} \in P_{(\mathrm{sat})}$) iff $(\forall n)$ $(\exists \mathbf{x} \in \mathbb{I}^n,\ \mathbf{x} \neq n * b)$ $(\forall \mathbf{y} \in \mathbb{I}^n)$ if $\mathbf{x} < \mathbf{y}$ then $\mathsf{F}(\mathbf{x}) = \mathsf{F}(\mathbf{y})$ and $(\exists \mathbf{z} \in \mathbb{I}^{n+1})$ $\mathsf{F}(\mathbf{z}) > \mathsf{F}(\mathbf{x})$.

4. S-statistics

In this section we present a particular family of aggregation operators, called S-statistics (or ordered conditional maximum). We then show which S-statistics are impact functions and in which cases the above-proposed properties are fulfilled.

For a different example of such analysis, we refer the reader to [4], where the class of L-statistics (or ordered linear combination), which generalize the ordered weighted averaging operator (OWA), is also considered.

Definition 10. A **triangle of coefficients** (compare [5,6]) is a sequence $\triangle = (c_{i,n} \in \mathbb{I} : i \in [n], n \in \mathbb{N})$.

Note that such object can be represented graphically by

$$\begin{array}{llll} c_{1,1} & & & \\ c_{1,2} & c_{2,2} & & \\ c_{1,3} & c_{2,3} & c_{3,3} & \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

Definition 11. The **S-statistic** associated with a triangle of coefficients $\triangle = (c_{i,n} \in \mathbb{I} : i \in [n], n \in \mathbb{N})$ is a function $\mathsf{S}_\triangle \in \mathcal{E}(\mathbb{I})$ such that

$$\mathsf{S}_\triangle(\mathbf{x}) = \bigvee_{i=1}^{n} c_{i,n} \wedge x_{(n-i+1)} \tag{5}$$

for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{I}^{1,2,\ldots}$, where $\vee$ and $\wedge$ denote the supremum (and hence the name) and infimum operators, respectively.

Without loss of generality we assume further on that $(\forall n)$ $c_{1,n} \leqslant c_{2,n} \leqslant \cdots \leqslant c_{n,n}$. Actually, because if

$$K_n = \left\{ k = 2, 3, \ldots, n :\ c_{k,n} \leqslant \bigvee_{i=1}^{k-1} c_{i,n} \right\}$$

then

$$\bigvee_{i=1}^{n} c_{i,n} \wedge x_{(n-i+1)} = \bigvee_{i=1, i \notin K_n}^{n} c_{i,n} \wedge x_{(n-i+1)}.$$

The Ordered Weighted Maximum operator (OWMax) defined for a fixed n, $\mathbb{I} = [0, 1]$ and $\triangle$ such that $\bigvee_{i=1}^{n} c_{i,n} = 1$ is an example of a function $\mathsf{S}_\triangle|_{\mathbb{I}^n}$. It was first introduced in [9].

This extension of the OWMax operator to arbitrary number of arguments is analogous to the construction in [6,25] done for OWA. For some general OWA operator weights construction methods see, e.g. [26,27].

The class of S-statistics was introduced in [4] as the ordered conditional maximum (OCM) operator. Some of its basic statistical properties were examined in [3]. For instance, it turns out that for random input data (continuous i.i.d. random variables) the distribution of an S-statistic is asymptotically normal.

It is worth noting that S-statistics generalize the well-known Hirsch h-index mainly used in the field of scientometrics. This index was originally defined in [2] for input vectors with elements in $\mathbb{N}_0$ as a function H such that

$$\mathsf{H}(x_1, \ldots, x_n) = \max\{i = 0, \ldots, n : x_{(n-i+1)} \geqslant i\},$$

where we assume that $x_{(n+1)} = x_{(n)}$. The h-index was proposed as a method of assessing the scientific merit of individual researchers by means of the number of citations received by their scientific papers. It quickly received much attention in the academic community [28,29]. Its popularity possibly arose from an appealing interpretation: the author of n papers has h-index of, say H, if H of his papers gained at least H, while the remaining $n-H$ papers — at most H citations. Interestingly, a similar object was defined earlier in the context of Bonferroni-type multiple significance testing (see, e.g. [30]).

For the expression of the h-index as a Sugeno integral (for more details compare [1]) of some function with respect to a fuzzy counting measure see [31]. The next proposition is inspired by this result. Let $\blacktriangle = (c_{i,n})_{i\in[n],n\in\mathbb{N}}$ be the triangle of coefficients which may be represented graphically as follows:

$$\begin{array}{llll} 1 & & & \\ 1 & 2 & & \\ 1 & 2 & 3 & \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

It may be easily shown (compare Theorem 14) that $\mathsf{S}_\blacktriangle$ is an impact function in $\mathbb{I} = [0,\infty]$.

Proposition 12. *Let $\mathbb{I} = [0,\infty]$. If $\blacktriangle = (c_{i,n})_{i\in[n],n\in\mathbb{N}}$ such that $c_{i,n} = i$ for $n \in \mathbb{N}$ and $i \in [n]$. Then for any $m \in \mathbb{N}, x_1,\ldots,x_m \in \mathbb{N}_0$,*

$$\mathsf{S}_\blacktriangle(x_1,\ldots,x_m) = \mathsf{H}(x_1,\ldots,x_m).$$

Proof. Let $H = \max\{i : x_{(n-i+1)} \geqslant i\}$. We have $\bigvee_{i=1}^{H} i \wedge x_{(n-i+1)} = H$ and $\bigvee_{i=H+1}^{n} i \wedge x_{(n-i+1)} = x_{(n-H)} < H+1$. However, since $x_{(n-H)} \in \mathbb{N}_0$, then $x_{(n-H)} \leqslant H$ and therefore $\bigvee_{i=1}^{n} i \wedge x_{(n-i+1)} = H$. $\square$

Note that we have $\mathsf{H}(2,1.5) = 1$ but $\mathsf{S}_\blacktriangle(2,1.5) = 1.5$. Generally, for arbitrary $\mathbf{x}$, $\mathsf{S}_\blacktriangle(\mathbf{x}) = \max\{H, x_{(n-H)}\} \in [H, H+1)$ or, on the other hand, $\mathsf{H}(\mathbf{x}) = \lfloor \mathsf{S}_\blacktriangle(\mathbf{x}) \rfloor$, where $\lfloor y \rfloor$ is the greatest integer $\leqslant y$. Therefore H is equivalent to $\mathsf{S}_\blacktriangle$ transformed by a particular non-decreasing function.

One may be interested for which triangles of coefficients the S operator is an aggregation function and/or impact function. The answer is given in Theorem 14. However, for its proof we need the following lemma.

Lemma 13. *Let $\mathbb{I} = [a,b]$ and $n \in \mathbb{N}$. Consider $\mathbf{c}, \mathbf{c}' \in \mathbb{I}^n$ such that $c_i \leqslant c_j$ and $c'_i \leqslant c'_j$ for all $i \leqslant j$. Then we have*

$$(\forall \mathbf{x} \in \mathbb{I}^n)\ \bigvee_{i=1}^{n} c_i \wedge x_{(n-i+1)} \geqslant \bigvee_{i=1}^{n} c'_i \wedge x_{(n-i+1)} \iff (\forall k \in [n])\ c_k \geqslant c'_k. \tag{6}$$

Proof. ($\Leftarrow$) Trivial.

($\Rightarrow$) Assume conversely. Let k be such that $c_k < c'_k$. Take any $\mathbf{x} \cong (k * c'_k, (n-k) * a)$. As $\mathbf{c}, \mathbf{c}'$ are nondecreasing, we have $\bigvee_{i=1}^{n} c_i \wedge x_{(n-i+1)} = c_k < \bigvee_{i=1}^{n} c'_i \wedge x_{(n-i+1)} = c'_k$, a contradiction. Therefore the proof is complete. $\square$

Additionally, it is easily seen that we have equality at the left side of (6) iff $(\forall k \in [n])\ c_k = c'_k$.

Theorem 14. *Let $\mathbb{I} = [a,b]$. Then for any $\triangle = (c_{i,n})_{i\in[n],n\in\mathbb{N}}$, $c_{i,n} \in \mathbb{I}$ such that $c_{i,n} \leqslant c_{j,n}$ for $i \leqslant j$ we have*

(a) $\mathsf{S}_\triangle \in \mathcal{E}_{\mathcal{A}}(\mathbb{I})$ *iff* $(\forall n)\ c_{n,n} = b$,
(b) $\mathsf{S}_\triangle \in \mathcal{E}_{\mathcal{I}}(\mathbb{I})$ *iff* $(\forall n)\ (\forall i \in [n])\ c_{i,n+1} \geqslant c_{i,n} \geqslant a$ *and* $\lim_{n\to\infty} c_{n,n} = b$.

We omit the simple proof, which bases on the previous lemma and the fact that for arbitrary $\triangle$ always $\mathsf{S}_\triangle \in P_{(I1)} \cap P_{(I5)}$. We now characterize the S-statistics fulfilling the properties given in the previous section.

Proposition 15. *For any $\mathbb{I} = [a,b]$ and any $\triangle = (c_{i,n})_{i\in[n],n\in\mathbb{N}}$, $c_{i,n} \leqslant c_{j,n}$ for $i \leqslant j$, such that $\mathsf{S}_\triangle \in \mathcal{E}_{\mathcal{I}}(\mathbb{I})$, the following holds*:

(a) $\mathsf{S}_\triangle \in P_{(a0)}$ *iff* $(\forall n)\ (\forall i \in [n]) c_{i,n+1} = c_{i,n}$.
(b) $\mathsf{S}_\triangle \in P_{(F0)}$ *iff* $\mathsf{S}_\triangle \in P_{(a0)}$.
(c) $\mathsf{S}_\triangle \in P_{(F+)}$ *iff* $(\forall n)$ *if* $c_{1,n} < b$ *then* $c_{1,n+1} > \bigvee_{i\in[n],c_{i,n}<b} c_{i,n}$.
(d) $\mathsf{S}_\triangle \in P_{(sat)}$ *iff* $(\forall n) c_{n,n} < c_{n+1,n+1} < b$.

Sketch of the proof.

(a) It follows from the remark to Lemma 13.
(b) Let us fix n. We should only show that $(\forall i \in [n])\ \bigvee_{j=1}^{i} c_{j,n} = \bigvee_{j=1}^{i} c_{j,n+1}$ implies $\mathsf{S}_\triangle \in P_{(s-)}$. Let $\mathsf{S}_\triangle(\mathbf{x}) = c_{j,n} \wedge x_{(n-j+1)}$ for some j. But as $\mathsf{S}_\triangle(\mathbf{x}) \leqslant x_{(n-j+1)}$ and $x_{(n-j+1)} \wedge \bigvee_{i=j+1}^{n+1} c_{i,n+1} \leqslant x_{(n-j+1)}$, it holds $\mathsf{S}_\triangle(\mathbf{x}, \mathsf{S}_\triangle(\mathbf{x})) = c_{j,n+1} \wedge x_{(n-j+2)} = c_{j,n} \wedge x_{(n-j+1)} = \mathsf{S}_\triangle(\mathbf{x})$.

(c) Let us fix n and let $c_{j,n} = \bigvee_{i\in[n],c_{i,n}<b} c_{i,n}$ for some $j \in [n]$. Take $\mathbf{x} = (n * c_{j,n})$. Then for any $\varepsilon > 0 \mathsf{S}_\triangle(\mathbf{x}, c_{j,n} + \varepsilon) > c_{j,n} = \mathsf{S}_\triangle(\mathbf{x})$ iff $c_{1,n+1} > c_{j,n}$. Now take any $\mathbf{y} \in \mathbb{I}^n$. Let $\mathsf{S}_\triangle(\mathbf{y}) = c_{j,n} \wedge y_{(n-j+1)}$ for some j. Then $c_{j,n} \wedge y_{(n-j+1)} < c_{1,n+1} \wedge ((c_{j,n} \wedge y_{(n-j+1)}) + \varepsilon)$.

(d) $\mathsf{S}_\triangle(n * c_{n,n}) = c_{n,n}$ and $(\forall \mathbf{y} \in \mathbb{I}^n : \mathbf{y} > (n * c_{n,n}))\ \mathsf{S}_\triangle(\mathbf{y}) = c_{n,n}$. There exists $\mathbf{z} \in \mathbb{I}^{n+1}$ such that $\mathsf{F}(\mathbf{z}) > \mathsf{F}(\mathbf{x})$ iff $b > c_{j,n+1} > c_{n,n}$ for some $j \in [n+1]$ ($\mathbf{z} = ((n+1) * c_{j,n+1})$). But $c_{i,n+1} \leqslant c_{j,n+1}$ for $i \leqslant j$ so it holds iff $c_{n+1,n+1} > c_{n,n}$, and the proof is complete. □

For example, the generalized h-index, $\mathsf{S}_\blacktriangle$, is clearly an F-insensitive and saturable impact function.

Moreover, the only $\mathsf{S}_\triangle \in \mathcal{E}_\mathcal{I}(\mathbb{I})$ which is both F-insensitive and $F+$ sensitive is equivalent to the sample maximum ($x_{(n)}$). This is because for F-insensitive S-statistics we have $c_{1,n} = c_{1,m}$ for all $n < m$. However, $F+$ sensitivity requires that when $c_{1,n} < b$, then it should increase at $c_{1,m}$. Therefore $c_{1,n} = b$ for any n. This indicates that the two properties together are too restrictive.

5. Quality assessment in the dynamic perspective

5.1. Motivation

In this section we discuss a problem strongly related to "the more the better" principle of the PAP. To be more illustrative let us focus our attention on the scientometric interpretation (Table 1A).

The problem is twofold. Firstly, perfect knowledge of the author's publications and number of their citations is assumed. In practice we gather bibliometric data from large databases, such as Thomson *Web of Science*, Elsevier *Scopus* or Google *Scholar* (see, e.g. [32,33]). The coverage of digital libraries is limited, so in most cases we are dealing with right-censored data (see [29] for discussion). In other words, the true number of citations received by a paper is at least equal to the value obtained from a given database.

Secondly, citations are merely an indicator of a paper's quality as actually perceived by the scientific community. Every paper, even the most influential one, has its initial number of citations equal to 0. That number grows as the work is recognized to be valuable (for some attempts to construct theories of citations in the social sciences see, e.g. [34–36]). Note that the acceptance of a paper for publication only expresses the fact this contribution meets at least some minimal journal standards.

Therefore, we may say that the process of publishing and citing in its very nature is **dynamic**, i.e. an input vector $\mathbf{x}$ shows only the current state of the producer and not its overall capability/potential. This state is very likely to change in the future as he publishes more works and his papers gain more citations. This process itself is always **accumulative** as the states $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$ in consecutive time intervals form a chain w.r.t. the relation $\trianglelefteq$, that is $\mathbf{x}^{(1)} \trianglelefteq \mathbf{x}^{(2)} \trianglelefteq \cdots$

Suppose someone described by $\mathbf{x}$ is applying for an academic tenure and he should determine his current impact $\mathsf{F}(\mathbf{x})$ for some $\mathsf{F} \in \mathcal{E}_\mathcal{I}(\mathbb{I})$. Can he be sure that his rating is not about to increase shortly, due to the fact that some new papers citing his work will have been published before his application will be considered?

On the other hand, in case of saturable impact functions, the real values of input vector coordinates are sometimes ignored. Consider $\mathsf{S}_\blacktriangle$ (discussed in Proposition 12) and $\mathbb{I} = [0, \infty]$. The function value of $\mathsf{S}_\blacktriangle(\mathbf{x}) = 4$ may be obtained for uncountably many $\mathbf{x} \in \mathbb{I}^4$: we have $\mathsf{S}_\blacktriangle(\mathbf{x}) = 4$ iff $x_{(n-3)} \geqslant 4$. However, it is not a sufficient condition for $\mathsf{S}_\blacktriangle(\mathbf{x}, 5) = 5$.

For another illustration let us consider the two vectors:

$$\mathbf{x} = (4,4,4,4,0,0,0,0), \quad \text{and}$$
$$\mathbf{y} = (9,8,7,6,4,4,3,1).$$

In both cases we have $\mathsf{S}_\blacktriangle(\mathbf{x}) = \mathsf{S}_\blacktriangle(\mathbf{y}) = 4$. However, not much is needed to increase $\mathsf{S}_\blacktriangle(\mathbf{y})$ even to 6 (actually, some increments of $y_{(3)}$ and $y_{(4)}$ or appropriate element additions are only necessary).

Generally, given an impact function $\mathsf{F} \in \mathcal{E}_\mathcal{I}(\mathbb{I})$ and a vector $\mathbf{x} \in \mathbb{I}^n$ we may try to raise the value of $\mathsf{F}(\mathbf{x}) < b$ either by

(o1) increasing the value of some of the elements of $\mathbf{x}$, i.e. finding $\mathbf{y} \in \mathbb{I}^n$ such that $\mathsf{F}(\mathbf{x} + \mathbf{y}) > \mathsf{F}(\mathbf{x})$, or
(o2) adding (concatenating) some elements to $\mathbf{x}$, i.e. finding $\mathbf{z} \in \mathbb{I}^m$ such that $\mathsf{F}(\mathbf{x}, \mathbf{z}) > \mathsf{F}(\mathbf{x})$, or
(o3) increasing the value of some of the elements and adding other elements simultaneously, i.e. finding $\mathbf{y}' \in \mathbb{I}^n$ and $\mathbf{z}' \in \mathbb{I}^m$ such that $\mathsf{F}(\mathbf{x} + \mathbf{y}', \mathbf{z}') > \mathsf{F}(\mathbf{x})$.

In each case some *effort/cost* (e.g. a publication of another paper, a marketing campaign) is required to improve the rating of the producer described by $\mathbf{x}$. However, one should be aware that there are impact functions and input vectors for which some of the above-mentioned actions do not result in the desired rating improvement. For example, in case of (o1), if F is saturable such $\mathbf{y}$ may sometimes not exist.

5.2. Effort-measurable functions

In this paragraph we distinguish a class of aggregation operators, called effort-measurable functions and its subclass, effort-dominable functions. Then we propose some general methods of creating possibility distributions that can help to describe how the aggregation operator's output value is likely to change when the above-mentioned operations (o1)–(o3) are applied to an input vector.

Let $\mathsf{F} \in P_{(I1)} \cap P_{(I4)} \cap P_{(I5)}$ and let $V_\mathsf{F} = \mathsf{F}[\mathbb{I}^{1,2,\ldots}]$ (the *image* of F). Moreover, for each $v \in V_\mathsf{F}$ let $\mathsf{F}^{-1}[v] := \{\mathbf{x} \in \mathbb{I}^{1,2,\ldots} : \mathsf{F}(\mathbf{x}) = v\}$ denote the *level set* of v. (Note that, technically, in the level set we should retain only the elements that are unique w.r.t. the relation $\cong$ to avoid ambiguities.)

Definition 16. We say that $\mathsf{F} \in P_{(I1)} \cap P_{(I4)} \cap P_{(I5)}$ is an **effort-measurable function** (denoted as $\mathsf{F} \in P_{(em)}$) iff for every $v \in V_\mathsf{F}$, $(\mathsf{F}^{-1}[v], \unlhd)$ is a partially ordered set with a least element.

In other words, $\mathsf{F} \in P_{(em)}$ iff $(\mathsf{F}^{-1}[v], \unlhd)$ is a bounded semi-lattice for all $v \in V_\mathsf{F}$.

Not all impact functions are effort-measurable. For example, some L-statistics, i.e. aggregation operators of the form $\mathsf{L}_\triangle(\mathbf{x}) = \sum_{i=1}^n c_{i,n} x_{(n-i+1)}$, such that $\mathsf{L}_\triangle \in \mathcal{E}_\mathcal{I}(\mathbb{I})$, do not belong to $P_{(em)}$.

Let us denote the least element of $\mathsf{F}^{-1}[v]$ by μ_v, i.e. $\boldsymbol{\mu}_v = \min \mathsf{F}^{-1}[v]$. Moreover, let M_F denote the set of all least elements corresponding to level sets for all v, i.e. $M_\mathsf{F} = \{\boldsymbol{\mu}_v : v \in V_\mathsf{F}\}$.

For example, consider the impact function $\mathsf{Max}(\mathbf{x}) = x_{(n)}$. We have $\mathsf{Max}^{-1}[v] = \{\mathbf{x} \in \mathbb{I}^{1,2,\ldots} : x_{(n)} = v\}$ and $\mu_v = \min \mathsf{Max}^{-1}[v] = (v)$.

Definition 17. We say that an effort-measurable function F is *effort-dominable* (denoted as $\mathsf{F} \in P_{(ed)}$) iff $(M_\mathsf{F}, \unlhd)$ is a chain.

Every effort-dominable function $\mathsf{F} \in P_{(ed)}$ may be defined as follows. For any $\mathbf{x} \in \mathbb{I}^{1,2,\ldots}$

$$\mathsf{F}(\mathbf{x}) = \max_{\mu_v \in M_\mathsf{F} : \mu_v \unlhd \mathbf{x}} v,$$

hence for $v, v' \in V_\mathsf{F}$ we have $\boldsymbol{\mu}_v \lhd \boldsymbol{\mu}_{v'} \iff v < v'$.

For example, the r_p-indices discussed in [37,38] are effort-dominable. Their definition will be recalled in Section 6, see (8).

Of course, not all effort-measurable functions are effort-dominable, e.g.

$$\mathsf{F}(\mathbf{x}) = a + (b-a)/5 \cdot \max\left\{i = 0, \ldots, \min\{n, 4\} : i = 0 \text{ or } x_{(n-i+1)} \geqslant 5 - i\right\}$$

is a counterexample ($\mathbb{I} = [a, b],\ \mathbf{x} \in \mathbb{I}^n$).

Now we turn back to our exemplary class of functions. We check which of the S-statistics are effort-dominable and effort measurable.

Proposition 18. *Let* $\triangle = (c_{i,n})_{i \in [n], n \in \mathbb{N}},\ c_{i,n} \leqslant c_{j,n}$ *for* $i \leqslant j$, *such that* $\mathsf{S}_\triangle \in \mathcal{E}_\mathcal{I}(\mathbb{I}),\ \mathbb{I} = [a, b]$. *Then the following conditions are equivalent*:

(a) *given any* n *and* $k = \min\{i : c_{i,n} = c_{n,n}\}$ *we have* $(\forall i < k)\ c_{i,n+1} = c_{i,n}$ *and* $(\forall j \geqslant k)\ c_{j,n+1} \geqslant c_{k,n}$,
(b) $\mathsf{S}_\triangle$ *is effort-dominable*,
(c) $\mathsf{S}_\triangle$ *is effort-measurable*.

Proof

(a ⇒ b) Let $v \in V_{\mathsf{S}_\triangle}$, n be the smallest number such that $c_{n,n} \geqslant v$. For any $m \geqslant n$,

$$\min\{\mathbf{x} \in \mathbb{I}^m : \mathsf{S}_\triangle(\mathbf{x}) = v\} \cong (k_m * v, (m - k_m) * a),$$

for some $k_m \in [m]$ (min is w.r.t. $\unlhd$). $c_{i,m} = c_{i,n}$ for $i < k_n$ and $c_{k_n,m} \geqslant c_{k_n,n}$ implies that $k_m = k_n$, so $\mu_v \cong (k_n * v, (n - k_n) * a)$ is the least element of $\mathsf{S}_\triangle^{-1}[v]$. However, given arbitrary $v' > v, v' \in V_{\mathsf{S}_\triangle}$, there exists $n' \geqslant n$ such that $(k_n * v, (n - k_n) * a) \lhd (k_{n'} * v', (n' - k_{n'}) * a) \cong \mu_{v'}$.

(b ⇒ c) By Definition 17.

(c ⇒ a) Please note that for any effort-measurable operator $\mathsf{S}_\triangle \in \mathcal{E}_\mathcal{I}(\mathbb{I})$ we have $\mu_v \cong (k * v, (n-k)*a)$, where $n = \min\{i : c_{i,i} \geqslant v\}$ and $k = \min\{i : c_{i,n} \geqslant v\}$. $(\forall m \geqslant n)\ (\forall i < k) c_{i,m} < v$, because otherwise $\mathsf{S}_\triangle(i * v, (m-i) * a) = v$ and $\mu_v \ntrianglelefteq (i * v, (m-i) * a)$ which leads to contradiction.

By Theorem 14, $c_{i,n} \leqslant c_{i,m} < v \leqslant c_{k,n}$. However, considering any $v \in V_{\mathsf{S}_\triangle}$, this statement implies $c_{i,n} = c_{i,m}$ for $i < \min\{j : c_{j,n} = c_{n,n}\}$ and the proof is complete. □

Moreover, this result implies that $\mathsf{S}_\triangle \in P_{(a0)} \Rightarrow \mathsf{S}_\triangle \in P_{(ed)}$.

5.3. Possibilistic approach

The modeling of complex processes underlying the changes of products' ratings can hardly be described by stochastic methods, even if we accept drastic simplifications and idealizations. Hence, we should try to develop approximate methods that can help to answer the question: Which input vectors are *more likely* to entail greater impact function values?

Thus our final goal is to establish a framework for some qualitative possibility relations between values of the impact function under study for different input data. Such possibility distribution functions might be useful for comparing the effects of increment and/or addition of elements on the aggregation operator's value.

Definition 19. An **impact possibility distribution** for an effort-measurable function $\mathsf{F} \in P_{(\mathrm{em})}$ is a mapping $\pi_{\mathsf{F}} : \mathbb{I}^{1,2,\ldots} \times \overline{\mathbb{R}} \to [0,1]$ such that for any $\mathbf{x} \in \mathbb{I}^{1,2,\ldots}$ the following conditions hold:

(a) $\pi_{\mathsf{F}}(\mathbf{x}; \mathsf{F}(\mathbf{x})) = 1$,
(b) if $v < \mathsf{F}(\mathbf{x})$ or $v \notin V_{\mathsf{F}}$ then $\pi_{\mathsf{F}}(\mathbf{x}; v) = 0$,
(c) if $\mathbf{x} \trianglelefteq \mathbf{y}$ and $v \geqslant \mathsf{F}(\mathbf{y})$, then $\pi_{\mathsf{F}}(\mathbf{x}; v) \leqslant \pi_{\mathsf{F}}(\mathbf{y}; v)$,
(d) if $v, v' \in V_{\mathsf{F}}$ are such that $\mathsf{F}(\mathbf{x}) \leqslant v < v'$ and $\mu_v \triangleleft \mu_{v'}$, then $\pi_{\mathsf{F}}(\mathbf{x}; v) \geqslant \pi_{\mathsf{F}}(\mathbf{x}; v')$.

The first condition states that the actual rating $\mathsf{F}(\mathbf{x})$ is obviously fully possible. Secondly, due to the accumulative nature of particular production-rating processes of concern, the values smaller that the actual one can not be obtained. The last two conditions ensure monotonicity with respect to input vectors and the relation $\trianglelefteq$ in (c), and, on the other hand, the output values of the impact function F and the standard ordering of reals in (d).

Further on we will write $\pi_{\mathsf{F},\mathbf{x}}(v)$ instead of $\pi_{\mathsf{F}}(\mathbf{x}; v)$ for convenience.

It might be seen easily that if $\mathsf{F} \in P_{(\mathrm{ed})}$ then $\pi_{\mathsf{F},\mathbf{x}}$ is non-increasing on $\mathrm{Supp}\,\pi_{\mathsf{F},\mathbf{x}} = \{v : \pi_{\mathsf{F},\mathbf{x}}(v) > 0\}$ because, by definition, we have $v < v' \Rightarrow \mu_v \triangleleft \mu_{v'} \Rightarrow \pi_{\mathsf{F},\mathbf{x}}(v) \geqslant \pi_{\mathsf{F},\mathbf{x}}(v')$.

Below we suggest two general methods for constructing impact possibility distributions.

5.4. Prediction based on effort metrics

The first method may be applied for effort-measurable functions. It is based on calculation of quasi-distances between an input vector $\mathbf{x}$ and the least elements of $\mathsf{F}^{-1}[v]$, $v \in V_{\mathsf{F}}$, i.e. vectors in M_{F}.

Definition 20. We say that a function $\mathsf{d} : \mathbb{I}^{1,2,\ldots} \times \mathbb{I}^{1,2,\ldots} \to \overline{\mathbb{R}}$ is an *effort metric* for an effort-measurable function $\mathsf{F} \in P_{(\mathrm{em})}$ if and only if for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{I}^{1,2,\ldots}$

(a) $\mathsf{d}(\mathbf{x},\mathbf{y}) \geqslant 0$ (non-negativity),
(b) if $\mathbf{y} \trianglelefteq \mathbf{z}$ then $\mathsf{d}(\mathbf{z},\mathbf{y}) = 0$ (left-to-right-flow),
(c) $\mathsf{d}(\mathbf{x},\mathbf{y}) \leqslant \mathsf{d}(\mathbf{x},\mathbf{z}) + \mathsf{d}(\mathbf{z},\mathbf{y})$ (triangle inequality).

Effort metrics may be used to express numerically the intuitive idea of the cost required to upgrade one input vector to another by means of the operations (o1)–(o3).

Each effort metric d is a quasi-metric. This is because it is non-negative, it satisfies the triangle inequality and $\mathsf{d}(\mathbf{x},\mathbf{x}) = 0$.

Proposition 21. *Given an effort metric* d *for* $\mathsf{F} \in P_{(\mathrm{em})}$ *and a non-increasing function* $\mathsf{g} : [0,\infty] \to [0,1]$ *such that* $\mathsf{g}(0) = 1$, *the following function*

$$\pi_{\mathsf{F},\mathbf{x}}(v) = \begin{cases} \mathsf{g}(\mathsf{d}(\mathbf{x}, \mu_v)) & \text{for } v \geqslant \mathsf{F}(\mathbf{x}) \text{ and } v \in V_{\mathsf{F}} \\ 0 & \text{otherwise} \end{cases}$$

is an impact possibility distribution for F.

Proof. Conditions (a) and (b) in Definition 19 obviously hold.

Consider $\mathbf{x} \trianglelefteq \mathbf{y}$ and any $v \geqslant \mathsf{F}(\mathbf{y})$, $v \in V_{\mathsf{F}}$. By Definition 20 we get $\mathsf{d}(\mathbf{y}, \mu_v) \leqslant \mathsf{d}(\mathbf{x}, \mu_v)$. Therefore $\pi_{\mathsf{F},\mathbf{x}}(v) \leqslant \pi_{\mathsf{F},\mathbf{y}}(v)$.

Now let $v, v' \in V_{\mathsf{F}}$ such that $\mathsf{F}(\mathbf{x}) \leqslant v < v'$ and $\mu_v \triangleleft \mu_{v'}$. Hence $\mathsf{d}(\mu_{v'}, \mu_v) = 0$ and consequently $\pi_{\mathsf{F},\mathbf{x}}(v) \geqslant \pi_{\mathsf{F},\mathbf{x}}(v')$. □

An example of prediction based on such construction will be considered in the next section.

5.5. Prediction based on exploring sets

The second approach utilizes a parameterized class $\{\mathsf{D}_q : q \in [0,1]\}$ that consists of effort-dominable functions, that is $\mathsf{D}_q \in P_{(\mathrm{ed})}$ for any $q \in [0,1]$. This very general method is applicable to impact functions which often ignore full information (e.g. saturable aggregation operators).

Definition 22. A set $\mathcal{D}_{\mathsf{F}} = \{\mathsf{D}_q \in P_{(\mathrm{ed})} : q \in [0,1]\}$ is called an **exploring set** for an effort-dominable function $\mathsf{F} \in P_{(\mathrm{ed})}$ iff

(a) $\mathsf{F} \in \mathcal{D}_{\mathsf{F}}$,
(b) $(\forall q \in [0,1]) V_{\mathsf{D}_q} = V_{\mathsf{F}}$,
(c) $(\forall v \in V_{\mathsf{F}})(\forall q, q' \in [0,1] : q \leqslant q') \min \mathsf{D}_q^{-1}[v] \trianglelefteq \min \mathsf{D}_{q'}^{-1}[v] \trianglelefteq \mathsf{F}^{-1}[v]$.

Intuitively, the functions in the exploring set require the same or even less *effort* to reach given impact rating v than F. Note that if $q \leqslant q'$ then $D_q(\mathbf{x}) \geqslant D_{q'}(\mathbf{x})$. This fact may be used to derive an immediate proof of the following proposition.

Proposition 23. *If* $\mathcal{D}_{\mathsf{F}} = \{\mathsf{D}_q : q \in [0,1]\}$ *is an exploring set for* $\mathsf{F} \in P_{(\mathrm{ed})}$, *and* $(\forall \mathbf{x})\ (\forall v \in V_{\mathsf{F}} \cap v \in [\mathsf{D}_1(\mathbf{x}), \mathsf{D}_0(\mathbf{x})])\ (\exists q \in [0,1])$ $\mathsf{D}_q(\mathbf{x}) = v$ *then*

$$\pi_{\mathsf{F},\mathbf{x}}(v) = \begin{cases} \max\{q \in [0,1] : \mathsf{D}_q(\mathbf{x}) = v\} & \text{for } v \in V_{\mathsf{F}} \cap [\mathsf{D}_1(\mathbf{x}), \mathsf{D}_0(\mathbf{x})] \\ 0 & \text{otherwise} \end{cases}$$

is an impact possibility distribution for F.

6. Examples

In this section we illustrate two prediction methods introduced above in a typical real-life PAP. Our data set consists of three vectors listed in Table 2 representing the number of citations of papers written by Polish computer scientists, Prof. A with $n_A = 19$ publications, Prof. B with $n_B = 29$ and Prof. C with $n_C = 18$ publications (the citation data were gathered from *Scopus* on January 20, 2009, and the names of the authors were intentionally masked).

Thus our PAP concerns scientists considered as producers and publications treated as their products (see Table 1A). Finally, the effort-dominable $\mathsf{S}_\blacktriangle$ operator discussed in Proposition 12, i.e. the generalized Hirsch index, is considered as a rating method. Assume that $\mathbb{I} = [0, \infty]$.

From the first glance at Table 2 it is clear that the outputs of our three authors differ. However, $\mathsf{S}_\blacktriangle(\mathbf{x}_A) = \mathsf{S}_\blacktriangle(\mathbf{x}_B) = \mathsf{S}_\blacktriangle(\mathbf{x}_C) = 7$ so our aggregation operator does not discriminate between them.

Note that the least element of $\mathsf{S}_\blacktriangle^{-1}[v]$, i.e. μ_v is equal to $(\lceil v \rceil * v)$ (compare the proof of Proposition 18), where $\lceil v \rceil = \min\{n \in \mathbb{Z} : n \geqslant v\}$ denotes the ceiling function. That is just seven items with rating 7 in the input vector $\mathbf{x}$ are sufficient to obtain $\mathsf{S}_\blacktriangle(\mathbf{x}) = 7$.

Below we present two examples. It should be stressed here that the choice of appropriate effort metrics or exploring sets is always dependent on the context. Some more formal approach to the topic deserves consideration in further research.

6.1. Example 1: Prediction based on an effort metric

Let us consider the following exemplary effort metric. For $\mathbf{y} \in \mathbb{I}^n$, $\mathbf{z} \in \mathbb{I}^m$ let

$$\mathsf{d}_*^2(\mathbf{y}, \mathbf{z}) = \begin{cases} \sum_{i=1}^{m} (0 \vee (z_{(m-i+1)} - y_{(n-i+1)}))^2 & \text{for } n \geqslant m, \\ \sum_{i=1}^{n} (0 \vee (z_{(m-i+1)} - y_{(n-i+1)}))^2 + \sum_{i=n+1}^{m} z^2_{(m-i+1)} & \text{for } n < m. \end{cases} \tag{7}$$

Also, let $\mathsf{g}_1(v) = (0 \vee 1 - v/22)$ (some non-increasing function). Thus we may obtain, by Proposition 21, an impact possibility distribution associated with d_*. Of course, the choice of d_* and g_1 here is somehow arbitrary. As we suggested above, the development of some general effort metrics construction methods is left for further research.

The impact possibility distributions for each of our three authors are given in Fig. 1. We may easily conclude that, in this case, the author B is more plausible to reach a greater impact function value than the others.

6.2. Example 2: Prediction based on an exploring set

Consider a family of r_p-indices proposed in [37]. The r_p-index for $p \geqslant 1$ is defined as follows:

$$r_p(\mathbf{x}) = \max\left\{r : (\forall i = 1, \ldots, \lceil r \rceil)(r^p - (i-1)^p)^{\frac{1}{p}} \leqslant x_{(n-i+1)}\right\}. \tag{8}$$

One can find its limit as $r_\infty \equiv \mathsf{S}_\blacktriangle$. It follows directly from the definition that it is an effort-dominable impact function for each p. Moreover (as shown in [37]), $r_p(\mathbf{x}) \leqslant r_q(\mathbf{x})$ for any $1 \leqslant q \leqslant p$.

This illustration is equivalent to the one given in [38].

It can be shown that, e.g. $\mathcal{D}_* = \{\mathsf{r}_{2-2^{1/q}} : q \in [0,1]\}$ is an exploring set for $\mathsf{S}_\blacktriangle$. It has a nice property that if $r_1(\mathbf{x}) = r_\infty(\mathbf{x})$ then the impact possibility distribution $\pi_{\mathsf{S}_\blacktriangle,\mathbf{x}}$ associated with $\mathcal{D}_*$ is a function given by (see [38] for the proof)

Table 2
State vectors representing the three authors.

$\mathbf{x}_A$	(103,20,16,16,10,9,8,5,4,4,4,3,2,2,2,1,0,0,0)
$\mathbf{x}_B$	(56,30,17,14,11,11,9,7,6,6,5,4,4,3,3,3,2,2,2,2,1,1,1,0,0,0,0,0,0)
$\mathbf{x}_C$	(39,34,23,17,16,7,7,5,2,2,1,1,0,0,0,0,0,0)

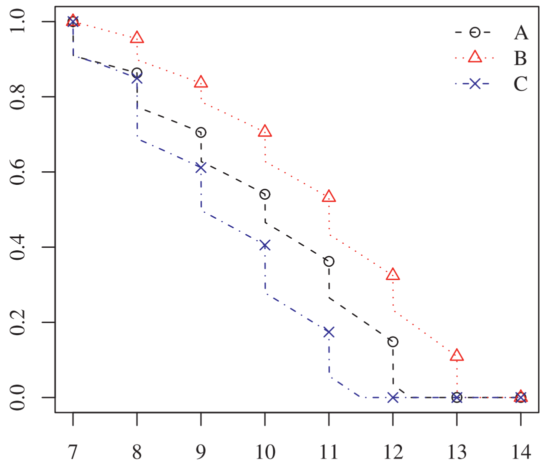

Fig. 1. Impact possibility distributions $\pi_{\mathsf{S}_\blacktriangle,\mathbf{x}_A}(v)$, $\pi_{\mathsf{S}_\blacktriangle,\mathbf{x}_B}(v)$, $\pi_{\mathsf{S}_\blacktriangle,\mathbf{x}_C}(v)$ associated with the effort metric d_* and the function g_1.

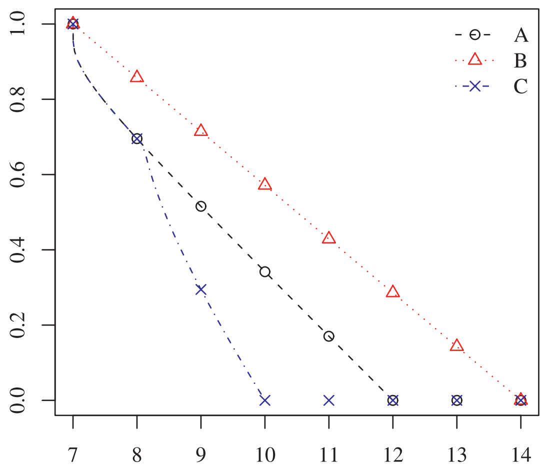

Fig. 2. Impact possibility distributions $\pi_{\mathsf{S}_\blacktriangle,\mathbf{x}_A}(v)$, $\pi_{\mathsf{S}_\blacktriangle,\mathbf{x}_B}(v)$, $\pi_{\mathsf{S}_\blacktriangle,\mathbf{x}_C}(v)$ associated with the exploring set $\mathcal{D}_*$.

$$\pi_{\mathsf{S}_\blacktriangle,\mathbf{x}}(v) = \begin{cases} 2 - \frac{1}{r_\infty(\mathbf{x})} v & \text{for } v \in [r_\infty(\mathbf{x}), 2r_\infty(\mathbf{x})], \\ 0 & \text{otherwise}. \end{cases} \tag{9}$$

The impact possibility distributions associated with the explorer set $\mathcal{D}_*$ for our three authors are given in Fig. 2. Please note that $\pi_{\mathsf{S}_\blacktriangle,\mathbf{x}_B}$ is of the form (9).

The conclusion here is similar to the one given for the previous example. Prof. *B* has the greatest possibility of increasing his impact rating. On the contrary, some papers of Prof. *C* were not cited so often and therefore the possibility that his score would increase considerably is not too high.

7. Conclusions

In this paper we discussed a family of aggregation operators, called impact functions, which may be regarded as a formal model of the Producer Assessment Problem. We examined some basic properties of impact functions to draw attention to the very nature of the PAP and the differences between them and the typical aggregation problems. We also considered a particular subfamily of impact functions, called S-statistics, which generalize the well-known *h*-index and OWMax operators.

After introductory considerations we focused on an important problem related to some instances of the PAP: the impreciseness and changeability of input data. We proposed a possibilistic approach which may help to make predictions on likely future input vector ratings. The suggested impact possibility distributions have an intuitive graphical interpretation and may be easily applied for comparing different state vectors.

All tools proposed in this paper may support the decision process and hence be useful in many areas, e.g. in scientometrics, marketing, management, etc. The proposed impact possibility distributions may for example be recommended as a complement to the Hirsch index.

However, many questions related to impact functions and the proposed prediction methods are still open. For example, the prediction based on the effort metrics may take into account different weights attributed to particular products, e.g. when it is much easier to increase the rating of a "good" product than of the "worse" one or when the costs required for getting a new product seem to be higher than those needed for the improvement of an existing one.

On the other hand, the prediction based on the exploring sets enables a precise requirement specification for each stage of the improvement process. It might be of special interest if the impact function under study takes a countable number of values only.

Of course, the choice of appropriate exploring sets or effort metrics is always dependent on the context or application of concern. However, some more formal, property-based approach to the topic deserves consideration in further research.

Moreover, we still lack some general S-statistics coefficient triangles construction methods. For related work on OWA and similar operators compare, e.g. [26,39,27,40].

Acknowledgments

The authors would like to express their gratitude to the Editor and the anonymous referees for their careful attention and precious suggestions to improve the manuscript.

References

[1] M. Grabisch, E. Pap, J.-L. Marichal, R. Mesiar, Aggregation Functions, Cambridge, 2009.
[2] J.E. Hirsch, An index to quantify individual's scientific research output, PNAS 102 (46) (2005) 16569–16572.
[3] M. Gagolewski, P. Grzegorzewski, S-statistics and their basic properties, in: C. Borgelt et al. (Eds.), Combining Soft Computing and Statistical Methods in Data Analysis, Springer-Verlag, 2010, pp. 281–288.
[4] M. Gagolewski, P. Grzegorzewski, Arity-monotonic extended aggregation operators, CCIS 80 (2010) 693–702.
[5] G. Mayor, T. Calvo, On extended aggregation functions, Proc. IFSA 1997, vol. 1, Academia, Prague, 1997, pp. 281–285.
[6] T. Calvo, A. Kolesarova, M. Komornikova, R. Mesiar, Aggregation operators: properties, classes and construction methods, in: T. Calvo, G. Mayor, R. Mesiar (Eds.), Aggregation Operators. New Trends and Applications, Studies in Fuzziness and Soft Computing, vol. 97, Physica-Verlag, New York, 2002, pp. 3–104.
[7] R.R. Yager, On generalized Bonferroni mean operators for multi-criteria aggregation, International Journal of Approximate Reasoning 50 (2009) 1279–1286.
[8] R.R. Yager, On ordered weighted averaging aggregation operators in multictriteria decision making, IEEE Transactions on Systems, Man, and Cybernetics 18 (1) (1988) 183–190.
[9] D. Dubois, H. Prade, C. Testemale, Weighted fuzzy pattern matching, Fuzzy Sets and Systems 28 (1988) 313–331.
[10] D. Dubois, H. Prade, Formal representations of uncertainty, in: D. Bouyssou, D. Dubois, M. Pirlot, H. Prade (Eds.), Decision-Making Process, ISTE, London, UK, 2009 (Chapter 3).
[11] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic. Theory and Applications, Prentice-Hall, PTR, New Jersey, 1995.
[12] S. Destercke, D. Dubois, E. Chojnacki, Unifying practical uncertainty representations. I: Generalized p-boxes, International Journal of Approximate Reasoning 49 (3) (2008) 649–664.
[13] S. Destercke, D. Dubois, E. Chojnacki, Unifying practical uncertainty representations. II: Clouds, International Journal of Approximate Reasoning 49 (3) (2008) 664–677.
[14] D. Dubois, H. Prade, P. Smets, A definition of subjective possibility, International Journal of Approximate Reasoning 48 (2) (2008) 352–364.
[15] G.J. Woeginger, An axiomatic analysis of Egghe's *g*-index, Journal of Informetrics 2 (4) (2008) 364–368.
[16] G.J. Woeginger, An axiomatic characterization of the Hirsch-index, Mathematical Social Sciences 56 (2) (2008) 224–232.
[17] T. Marchant, An axiomatic characterization of the ranking based on the *h*-index and some other bibliometric rankings of authors, Scientometrics 80 (2) (2009) 325–342.
[18] T. Marchant, Score-based bibliometric rankings of authors, Journal of the American Society for Information Science and Technology 60 (6) (2009) 1132–1137.
[19] I. Palacios-Huerta, O. Volij, The measurement of intellectual influence, Econometrica 72 (3) (2004) 963–977.
[20] A. Quesada, More axiomatics for the Hirsch index, Scientometrics 82 (2010) 413–418.
[21] B. de Finetti, Sul significato soggettivo della probabilitá, Fundamenta Mathematicae 17 (1931) 298–329.
[22] J. von Neumann, O. Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, Princeton, 1947.
[23] K.J. Arrow, A difficulty in the concept of social welfare, Journal of Political Economy 58 (4) (1950) 328–346.
[24] K.O. May, A set of independent necessary and sufficient conditions for simple majority decision, Econometrica 20 (4) (1952) 680–684.
[25] T. Calvo, G. Mayor, Remarks on two types of extended aggregation functions, Tatra Mountains Mathematical Publications 16 (1999) 235–253.
[26] B.S. Ahn, Parameterized OWA operator weights: an extreme point approach, International Journal of Approximate Reasoning 51 (7) (2010) 820–831.
[27] T. Calvo, G. Mayor, J. Torrens, J. Suner, M. Mas, M. Carbonell, Generation of weighting triangles associated with aggregation functions, International Journal of Uncertainty, Fuzziness and Knowledge-based Systems 8 (4) (2000) 417–451.
[28] P. Ball, Index aims for fair ranking of scientists, Nature 436 (2005) 900.
[29] S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera, *h*-index: a review focused on its variants, computation and standardization for different scientific fields, Journal of Informetrics 3 (2009) 273–289.
[30] Y. Benjamini, Y. Hochberg, Controlling false discovery rate: a practical and powerful approach to multiple testing, Journal of the Royal Statistical Society. Series B 57 (1) (1995) 289–300.
[31] V. Torra, Y. Narukawa, The *h*-index and the number of citations: two fuzzy integrals, IEEE Transactions on Fuzzy Systems 16 (3) (2008) 795–797.
[32] M. Franceschet, A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar, Scientometrics 83 (1) (2010) 243–258.
[33] E.S. Vieira, J.A.N.F. Gomes, A comparison of Scopus and Web of Science for a typical university, Scientometrics 81 (2) (2009) 587–600.
[34] E. Garfield, Can citation indexing be automated? in: M.E. Stevens, V.E. Giuliano, L.B. Heilprin (Eds.), Proceedings of the Statistical Association Methods for Mechanized Documentation, Washington, 1964, pp. 189–192.
[35] L. Bornmann, H.-D. Daniel, What do citation counts measure? A review of studies on citing behavior, Journal of Documentation 64 (1) (2008) 45–80.
[36] P.M. Davis, Reward or persuation? The battle to define the meaning of a citation, Learned Publishing 21 (2009) 5–11.

[37] M. Gagolewski, P. Grzegorzewski, A geometric approach to the construction of scientific impact indices, Scientometrics 81 (3) (2009) 617–634.
[38] M. Gagolewski, P. Grzegorzewski, Possible and necessary *h*-indices, in: Proceedings of the IFSA/Eusflat, 2009, pp. 1691–1695.
[39] B.S. Ahn, Preference relation approach for obtaining OWA operators weights, International Journal of Approximate Reasoning 47 (2) (2008) 166–178.
[40] R.R. Yager, Prioritized aggregation operators, International Journal of Approximate Reasoning 48 (1) (2008) 263–274.